

## EPIWORK WP3 Third Year Report

### 1. Publishable summary

#### *Project objectives for the period*

*Please provide an overview of the project objectives for the reporting period in question, as included in Annex I of the Grant Agreement.*

*Please include a summary of the recommendations from the previous reviews (if any) and indicate how these have been taken into account.*

We are developing an information platform to mediate access to distributed collections of public health data, offering an easy and safe way to share data for those data providers who want to collaborate with epidemiological modellers. Researchers will use this platform in multiple ways: i) as catalogue of data sources containing the metadata describing existing databases; ii) as a forum to publish information about their own data, seeking modellers to collaborate with, and/or to seek sources of data that could be of interest to their epidemiological modelling efforts; and finally, iii) as the host of mediating software that can automatically process queries for epidemiological data available from the information sources connected to the platform.

Work on the second year of the project was focused on the prototyping of the infrastructure for the platform, characterizing epidemic data and developing a common meta-model for querying data of interest for epidemiological modelling. The platform mediates access to distributed collections of public health data, offering an easy and safe way to share data for those data providers who want to collaborate with epidemiological modellers through a publicly available Web interface.

Research work on the third year of the project was focused on setting the access control framework and researching ontologies for interlinking epidemic data using semantic web standards. In parallel, we continued the development of a second version of the Epidemic Marketplace with much improved usability for curating epidemics data. The second version is now running on the Epidemic Marketplace website at <http://epimarketplace.net>.

Research work on the forth year of the project will be centred on the development of methods for assisting collaborators in easily find each other, and define mutual voluntary agreements for sharing their data, along with intensified work on populating the Epidemic Marketplace with epidemic data using the information collection tools and services developed by the consortium in previous years.

### 2.1 Background Information on WP3

Work package number	3	Start date or starting event:	Month 1
Work package title	Information platform		
Activity type	RTD		

<b>Participant number</b>	12	1	2	3	4	5	9	10
<b>Participant short name</b>	FFCUL	ISI	FGC-IGC	TAU	MPI-DS	AIBV	BIU	FBK-IRST
<b>Person-months per participant</b>	82	60	4	8	42	2	11	6

This Work Package is lead by FFCUL. The whole WP3 activity is structured into four tasks:

### **Task 3.1 – Data Collection.**

**Participants:** FFCUL, ISI, FBK-IRST, BIU, MPI-DS, FGC-IGC, AIBV.

**Description:** Realistic simulations of epidemic processes crucially depend on the availability of datasets describing human behaviour and pathogen-host interactions. Datasets include population movement data, social and behavioural data, health related data, geographic data, detailed geo-temporal epidemic incidence and immunization data, pathogen evolution and multi-strains circulation data. Data can come from a variety of different sources, including hospital records, country statistics, Web content, and others. It can range from a global scale, such as the worldwide air transportation infrastructure, down to the detailed description of individual **activities** at a minute-by-minute scale. This task will create a catalogue of databases of epidemiological data across Europe, with extensive meta-data describing the main characteristics of the available information sources. This catalogue will be integrated with a collaborative platform that will be set up for online discussion and exchange of meta-data among the participants.

### **Task 3.2 – Meta-Model Design.**

**Participants:** FFCUL, IGC, ISI.

**Description:** While some of the previously mentioned datasets are freely available on the Web (e.g. WHO Global Health Atlas, Eurostat), they are often scattered in different repositories, cover partial regions of the world and come in different formats, according to different standards and classifications. The project envisions a unified and integrated approach for the management of these resources, with the design and implementation of an Epidemic Marketplace Platform, publicly available on the web. The platform supports the sharing and management of epidemic datasets and resources as well as their rating, annotation, and selection. It is an on-line social networking site that will serve researchers, practitioners, and educators all over the world to foster a virtual community for epidemic research. It will support the exchange of resources as well as user interactions. Based on a Web2.0 approach, users will become active participants, sharing information and data, and

collaborating online, rather than being satisfied with a passive information consumer/viewer role. We envision proposing a simple reference format, which will facilitate the navigation and use of the datasets. Each dataset will come with a metadata file, signalling general metadata for resource management, containing data such as: the title, the date of submission, version, the source of the data and coverage. Moreover, the metadata will include information for a more thorough description of the data included in the dataset, providing a framework for a more specific description, for example, of epidemiologic and geographic data. The Marketplace will support flexible and intuitive tools for navigation and selection of resources. Standard classifications as well as tagging systems proposed by users will be supported.

### **Task 3.3 – Epidemic Marketplace Platform.**

**Participants:** FFCUL, ISI.

**Description:** This task will implement a platform based on the integration of grid technology and publicly available services and software on the web to support the sharing and management of epidemic datasets and resources as well as their rating, annotation, and selection. The Epidemic Marketplace Platform will be an on-line social networking site that will serve researchers, practitioners, and educators all over the world to foster a virtual community for epidemic research. It will support the exchange of resources as well as user interactions. Based on some of the Web2.0 characteristics, users will become active participants, generating information and providing data for sharing, and collaborating online, rather than being satisfied with a passive information consumer/viewer role. More specifically, researchers can use and contribute to the Marketplace in several different ways. They can: (1) use it as a catalogue of data sources containing the metadata describing existing databases; (2) view, download, tag, and comment on the available resources; (3) provide compliant datasets and relevant information; (4) use it as a forum where to publish information about their own data, seek modellers to collaborate with, share and distribute their new findings.

### **Task 3.4 – Evaluation and monitoring of the use of the catalogue and collaboration services.**

**Participant:** FFCUL.

**Description:** This task involves the monitoring of epidemiological data exchanges performed through the mediating services platform. The evaluation will assess not only the coverage of the catalogued resources, but the users' satisfaction with the user interface and integrated collaborative tools made available through the epidemiological marketplace platform. More importantly, the analysis of the collected datasets and their annotations and

usage will provide a rich environment for deriving an epidemiology ontology, which will help further on the integration and communication among the community of epidemiologists.

## **2. WP Work progress and achievements during the period**

- *A summary of progress towards objectives and details for each task;*
- *Highlight clearly significant results;*
- *Team publications within the cope of the project (please provide a pdf of the publication if possible);*
- *List of outreach activities (conferences, Invited talks, presentations, workshops, tutorials,);*
- *List of press releases or media coverage, any particular dissemination activity;*
- *unanticipated finding, opportunity etc.*
- *If applicable, explain the reasons for deviations from Annex I and their impact on other tasks as well as on available resources and planning;*
- *If applicable, explain the reasons for failing to achieve critical objectives and/or not being on schedule and explain the impact on other tasks as well as on available resources and planning (the explanations should be coherent with the declaration by the project coordinator) ;*
- *a statement on the use of resources, in particular highlighting and explaining deviations between actual and planned person-months per work package and per beneficiary in Annex I (Description of Work)*
- *If applicable, propose corrective actions.*

## 2.2 *Progress in the Reporting Period*

1. Mário Silva and Dulce Domingos attended the Epiwork Review in Brussels in March, 2011 – Mário Silva presented Deliverable D3.3 to the EC and gave an overview of the progress and challenges in WP3. The demo of the first functional Epidemic Marketplace prototype with all the anticipated components was presented to the reviewers.
2. **[opportunity]** The LASIGE team established contacts with other teams involved in projects for the development of biomedical ontologies and terminologies. Miquel Porta, author of the popular “Dictionary of Epidemiology” is very interested in our development of an epidemiological meta-model to describe epidemiological datasets and a possible collaboration may happen.
3. Completed evaluation of Integration of Drupal with Fedora Commons, using the Islandora module extension to Drupal and using this software as basis for a new version of the Repository/Forum of the Epidemic Marketplace and decided for a reorganization of the software architecture of the frontend, which will now interface with Fedora Commons through the EM webservice.
4. Complete re-write of the EM web-services, which do not require the Muradora software and support extended search services based on Apache Solr.
5. Corrado Gioanini visited LASIGE for one week in May for planning the integration of the GLEaMviz platform and the Epidemic Marketplace. Jointly developed the basic use-cases involved in this integration.
6. **[outreach]** The LASIGE team published and presented two full papers at the International Conference of Biomedical Ontologies, in July, which was a significant step towards the disclosure of our work to this important community.
7. **[outreach]** Mário J. Silva was invited to participate in a FuturICT Ethics meeting at ETH Zurich, where a group is being formed under the FuturICT FET Flagship initiative to address the issues of privacy and ethics in IT systems manipulating societal data.

8. **[outreach]** Francisco Couto was invited to give a seminar at EBI in April 2011 on biomedical ontologies.
9. Continued design of new access control method to be implemented in the EM.
10. The second version of the Epidemic Marketplace software, EM 2.0, platform has been in the operation since the end of 2011. Major features: new user interface and component integration based on the Drupal Content-Management System, improved access control management, redesigned web-services.
11. **[outreach]** The LASIGE team participated with five of its members in the EE2, Epiwork/Epifor Workshop, with three accepted posters.

#### **Publications and Presentations:**

1. **[presentation]** F. Couto, Exploring the semantics of biomedical ontologies, in External Seminar at European Bioinformatics Institute, Hinxton, UK, April 2011.
2. **[presentation]** Mário J. Silva. Privacy and Crowdsensing: Can't we just be friends? FuturICT's Ethics Meeting, Zurich, June 2011.
3. **[publication + presentation]** Catia Pesquita, Francisco Couto, Where GO is going and what it means for ontology extension. Proceedings of ICBO 2011, International Conference on Biomedical Ontology, July 2011.
4. **[publication + presentation]** Bruno Tavares, Hugo Bastos, Daniel Faria, Joao D. Ferreira, Tiago Grego, Catia Pesquita, Francisco Couto, The Biomedical Ontology Applications (BOA) framework. Proceedings of ICBO 2011, International Conference on Biomedical Ontology, July 2011.
5. **[publication + presentation]** João D. Ferreira, Francisco Couto, Generic semantic relatedness measure for biomedical ontologies. Proceedings of ICBO 2011, International Conference on Biomedical Ontology, July 2011.
6. **[presentation]** Presentation of EPIWORK to students of the Master/PhD in Epidemiology at the Faculty of Medicine of the University of Lisbon, in an invited 3 hours seminar in July 2011, by Mário J. Silva.

7. **[publication]** Francisco Couto, Mário J. Silva. Disjunctive Shared Information between Ontology Concepts: application to Gene Ontology. Journal of Biomedical Semantics. Vol 2(5). doi:10.1186/2041-1480-2-5.
8. **[presentation]** Mário J. Silva. Challenges in Societal Data Management. Keynote presentation at IDEAS'11: 15th International Database Engineering & Applications Symposium. Lisbon, September, 2011.
9. **[publication + presentation]** Juliana Duque, Mediação Dados-Informação: Design de Informação para a Epidemic Marketplace Master Thesis, University of Lisbon, School of Fine Arts, November 2011 (in Portuguese).
10. **[publication + presentation]** Carla Patrícia Freitas Sousa, Epidemic Marketplace: Repositório e Web Services, Master Thesis, University of Lisbon, Faculty of Sciences, January 2012.
11. **[publication]** Zamite, J., Silva, F., Couto, F., Silva, M. 2011: MEDCollector: Multisource epidemic data collector. Transactions on Large-scale Data-and Knowledge-centered Systems IV: Special Issue on Database Systems for Biomedical Applications (6990), pp. 40-72. Springer-Verlag New York, Inc. ISBN 978-3-642-23739-3
12. **[presentation]** Corrado Gioannini, João Zamite, Integrating the Gleanviz Simulator Tool with the Epidemic Marketplace Platform. Poster presented at EE2, Epiwork/Epifor 2nd International Workshop: Facing the Challenge of Infectious Diseases. 2012.
13. **[presentation]** João Zamite, Dulce Domingos, Mário J. Silva, Owner-Centred Group-Based Access Control for Epidemic Resources. Poster presented at EE2, Epiwork/Epifor 2nd International Workshop: Facing the Challenge of Infectious Diseases. 2012. [http://xldb.di.fc.ul.pt/xldb/publications/Zamite.etal:EpidemicGroupBasedAccessControl:2012\\_poster.pdf](http://xldb.di.fc.ul.pt/xldb/publications/Zamite.etal:EpidemicGroupBasedAccessControl:2012_poster.pdf)
14. **[presentation]** João D. Ferreira, Francisco M. Couto, Mário J. Silva, Ontologies in the Epidemiological Domain. Poster presented at EE2, Epiwork/Epifor 2nd International Workshop: Facing the Challenge of Infectious Diseases. 2012. [http://xldb.di.fc.ul.pt/xldb/publications/Ferreira.etal:OntologiesInThe:2012\\_poster.pdf](http://xldb.di.fc.ul.pt/xldb/publications/Ferreira.etal:OntologiesInThe:2012_poster.pdf)

15. **[publication]** João D. Ferreira, Catia Pesquita, Francisco Couto, Mário J. Silva. Epiwork Deliverable 3.5: Epidemic Data Ontology. Technical Report. University of Lisbon, Faculty of Sciences, LASIGE, January 2012.
16. Mário J. Silva, Francisco M. Couto, Dulce Domingos, João Ferreira, Paulo Graça, Tiago Posse, Carlos Sousa, João Zamite. Epiwork Deliverable 3.4 Epidemic Marketplace Platform Report at Month 36, Technical Report. University of Lisbon, Faculty of Sciences, LASIGE, January 2012.
17. **[presentation]** F. Couto, Untangling Biomedical Ontologies, Practical workshop: Bioinformatics and Systems Modelling, Faculty of Sciences, University of Lisbon, 2011.

#### **Activities in the third year of the project in Task 3.1:**

- The data collection activity started at the end of the first year and continued as before, with datasets being periodically assembled and uploaded into the Epidemic Marketplace.
- With the release of version 2.0 of the Epidemic Marketplace platform, all the datasets and metadata, which had been loaded in the initial prototype have been converted to the updated meta-data format and migrated to the new platform.

#### **Activities in the third year of the project in Task 3.2:**

- In the period, we worked on the development of a proposal of a network of ontologies for representing the data most commonly used in epidemiology, epidemics modelling in particular, reported in the **Deliverable D3.5**, which was completed by the end of the third year.

#### **Activities in the third year of the project in Task 3.3:**

- The second prototype of the Epidemic Marketplace was presented in the March 2011 Project Review in Brussels.
- WP3/WP4 collaboration on the integration of the computational platform with the data platform.



- Public release of the Epidemic Marketplace at <http://epimarketplace.net>. Anyone from the community can now define collections and retrieve/upload datasets.
- Continued design of new access control system to be implemented in the EM.
- EM 2.0 became the production version at the official site <http://epimarketplace.net>. Major features: new user interface and component integration based on the Drupal Content-Management System, improved access control management, redesigned web-services.
- Integration of GleanViz with the Epidemic Marketplace. We jointly developed an interface that will enable GleanViz upload simulation results as datasets to the EM platform to be shared by the community.

#### **Activities in the third year of the project in Task 3.4:**

- Continued work on planning of the monitoring and log data collection and analysis tasks. Progress on this domain is documented in Deliverable D3.4.
- Initiated planning on how to use the resources, especially their meta-data, for inferring relationships among the ontological concepts (this is addressed in Deliverable D3.5).

## **2.3 Outcast for WP3**

Focus for the 4th year will be on consolidating and integrating with the EPIWORK partner's systems and tools, in particular the Computational Platform. In addition, we will be focusing on enriching the Epidemic Marketplace with more datasets and semantifying with NERO part of the resources.

In parallel, we will continue developing the software of the Epidemic Marketplace, improving its interfaces and populating it with epidemic datasets.

The envisaged activities will mainly involve:

- Enrichment of the EM catalogue. This will entail not only the addition of more resources to the catalogue, but also the automatic insertion of rich annotations using the NERO ontologic concepts.
- Implementation of new ideas on social network access control in EM 2.0 and provide more support for locating modelers and data providers and mediating the negotiation of access to resources.
- Implementation of support for resource collections management, which will be essential for scaling the platform for handling large sets of resources.
- Tightening the integration with Computational Platform, supporting the sharing of GLEaMviz simulations through the Epidemic Marketplace, and demonstrating distributed operation with one EM node in the cloud.
- Providing tools for selecting datasets or parts of the information thereof from the EM collection for retrieval access by epidemic modelling tools. Development of the first demonstrations of information search based on ontologic knowledge integrated in the Epidemic Marketplace.